

# **Comments of the Internet Archive on the U.S. Copyright Office Notice of Inquiry on the Digital Millennium Copyright Act Section 512 Safe Harbors**

**February 21, 2017**

## **Introduction**

The Internet Archive submitted comments in response to the Copyright Office's Notice of Inquiry regarding the DMCA Section 512 Safe Harbor system on March 22, 2016. A representative of the Internet Archive participated in the round table discussion in San Francisco in May of 2016. We welcome this opportunity to continue the conversation and provide our responses to the Copyright Office's request for additional comments.

**1. As noted above, there is great diversity among the categories of content creators and ISPs who comprise the Internet ecosystem. How should any improvements in the DMCA safe harbor system account for these differences? For example, should any potential new measures, such as filtering or stay- down, relate to the size of the ISP or volume of online material hosted by it? If so, how? Should efforts to improve the accuracy of notices and counter- notices take into account differences between individual senders and automated systems? If so, how?**

The DMCA safe harbor system should apply to all ISPs equally. One of the hallmarks of the Internet is that it is a level playing field, where all comers play by the same set of rules. This means that even small players can compete against the big incumbents. This enables more competition among platforms, and more choices for consumers.

But for the Internet to remain a level playing field, all participants must play by the same set of rules. Applying one set of rules to "small" players and another to "big" players would not only undermine the fundamental values of the open web and create unpredictable market distortions, but it would also be extremely difficult to do from a practical standpoint.

On the Internet, it is possible to be both "big" and "small" at the same time. For example, in terms of popularity and amount of material hosted, the Internet Archive is a relatively "big" player on the Internet. Ranked among the 250 most popular sites, millions of people visit the Internet Archive each day. Yet, the Internet Archive is a small non-profit organization that employs only 125 people. By way of comparison, Uber.com

is ranked as the 325 most popular website, but it has a \$62.5 billion dollar valuation<sup>1</sup> and employs between 5,000 and 10,000 people.<sup>2</sup>

Moreover, the Internet Archive hosts a huge amount of data -- 30 petabytes as of the end of 2016. Many of the large hosting companies do not disclose the size of their libraries, or the volume of content they host. It could be very difficult for regulators to get this information in order to determine which companies might be subject to the regulation.

Further, the volume of content hosted by a platform may bear little or no relationship to the number of DMCA notices that platform receives. By way of example, the vast majority of the content hosted on the Internet Archive is not subject to the notice and takedown regime at all because it is not posted at the direction of users, for example the contents of the Wayback Machine and the Political TV Ad Archive. Of the remaining materials, many are in the public domain or posted with permission.<sup>3</sup> As such, only a very small proportion of material hosted on the Internet Archive might even be relevant to the DMCA safe harbors. Yet if measured by overall amount of data hosted, then the Internet Archive would likely be counted as "big" and therefore subject to additional regulations/obligations.

Moreover, regulatory parsing of ISPs in this manner seems unnecessary. Some of the largest platforms have already voluntarily developed systems that go well beyond the requirements of the statute--YouTube's Content ID system is the prime example. It is far preferable to encourage large platforms who receive a high volume of DMCA takedown notices to develop systems that work for their specific platforms than to technically mandate a specific technical response that may make no sense for the majority of platforms.

It is much more straightforward from a regulatory perspective to put differing requirements on those using automated software to send DMCA notices. Automated systems are easy to identify, and should be required to go through a technical audit to ensure that they have an acceptably low error rate to reduce the overall number of false/incomplete/inadequate DMCA notices sent.

## **2. Several commenters noted the importance of taking into account the perspectives and interests of individual Internet users when considering any changes to the operation of the DMCA safe harbors. Are there specific issues for**

---

<sup>1</sup> Eric Newcomer, *Uber Raises Funding at \$62.5 Billion Valuation*, Bloomberg Technology, (December 3, 2015) <https://www.bloomberg.com/news/articles/2015-12-03/uber-raises-funding-at-62-5-valuation>

<sup>2</sup> Crunchbase, <https://www.crunchbase.com/organization/uber#/entity>

<sup>3</sup> See for example, the Live Music Archive: <https://archive.org/details/etree>

**which it is particularly important to consult with or take into account the perspective of individual users and the general public? What are their interests, and how should these interests be factored into the operation of section 512?**

The public interest should always be taken into account. The general Internet-using public has an interest in a balanced copyright system that allows a diversity of platforms to grow and thrive.

**3. Participants expressed widely divergent views on the overall effectiveness of the DMCA safe harbor system. How should the divergence in views be considered by policy makers? Is there a neutral way to measure how effective the DMCA safe harbor regime has been in achieving Congress' twin goals of supporting the growth of the Internet while addressing the problem of online piracy?**

Policy makers must keep in mind that Section 512 does not exist in a regulatory vacuum. Any reform of Section 512 of the DMCA should be undertaken only in light of the entire copyright system, including consideration of the exceedingly high statutory damages rightsholders can claim, as well as the Section 1201 anti-circumvention provisions which primarily benefit rightsholders at the expense of technology platforms and consumers.

**4. Several public comments and roundtable participants noted practical barriers to effective use of the notice- and-takedown and counter-notice processes, such as differences in the web forms used by ISPs to receive notices or adoption by ISPs of additional requirements not imposed under the DMCA (e.g., submission of a copyright registration or creation of certain web accounts). What are the most significant practical barriers to use of the notice-and-takedown and counter-notice processes, and how can those barriers best be addressed (e.g., incentives for ISPs to use a standardized notice/counter-notice form, etc.)?**

The Internet Archive does not believe that a standardized web form would be helpful in terms of mitigating the types of errors we see on a regular basis. We accept notices in many different forms, and this works well for rightsholders and for our methods of human review. We are concerned that a standard web form might force certain types of automation that might take humans out of the loop in determining the validity of takedown requests. As we noted in our March 22, 2016 comments, the Internet Archive regularly receives incomplete, inaccurate, and false DMCA takedown requests. In order to ensure that huge swaths of legitimate materials are not removed in error, each and every notice is reviewed by a real person. Although this is burdensome, we far prefer it to a system where more automation would result in fewer safeguards against the removal of legitimate content.

**6. Participants also noted disincentives to filing both notices and counter-notices, such as safety and privacy concerns, intimidating language, or potential legal costs. How do these concerns affect use of the notice-and-takedown and counter-notice processes, and how can these disincentives best be addressed?**

It should be possible to file a counter-notice through an agent, just as it is possible to file a takedown notice through an agent. This would allow Internet users to remain anonymous rather than revealing their identity based only on an accusation of copyright infringement, which can be sent for malicious or abusive reasons. Section 512(h) of the DMCA allows rightsholders to obtain the identity of alleged infringers should they require it, so this change would not deprive rightsholders of the ability to bring legitimate lawsuits. But it might deter the use of DMCA takedown notices for the purposes of harassment.

**7. Some participants recommended that the penalties under section 512 for filing false or abusive notices or counter-notices be strengthened. How could such penalties be strengthened? Would the benefits of such a change outweigh the risk of dissuading notices or counter-notices that might be socially beneficial?**

Section 512(f) is completely toothless. Example after example shows that there is no remedy for false or abusive takedown notices. The *Lenz v. UMG* case has been in ongoing litigation for 10 years. The average Internet user simply cannot count on a decade of *pro bono* legal assistance from the electronic Frontier Foundation to get relief. Many other attempts have failed. In just the last few months, two more courts rejected 512(f) claims: *Ouellette v. Viacom International, Inc.*, 2016 WL 7407244 (9th Cir. Dec. 22, 2016) and *Opinion Corp. v. Roca Labs, Inc.*, 2016 WL 6824383 (M.D. Fla. Nov. 17, 2016).

The subjective bad faith requirement created by the *Rossi v. MPAA* case has made it impossible to win remedies under 512(f) and to thereby deter bad notices. One way to create real penalties for abusive takedown notices would be to include automatic statutory damages under 512(f), to mirror the statutory damages for copyright infringement. As under Section 504, there could be a range starting at \$750 per false notice, with an upper range for egregious and intentionally speech-chilling behavior. This would likely incentivise senders to ensure their notices are more accurate and not sent simply to censor or harass a user.

There should also be more incentives for ISPs to take the risk of rejecting bad notices. For example, statutory damages could be remitted to \$0 if the ISP has a reasonable belief that the manner in which the material is used is fair use.

**9. Many participants supported increasing education about copyright law generally, and/or the DMCA safe harbor system specifically, as a non- legislative**

**way to improve the functioning of section 512. What types of educational resources would improve the functioning of section 512? What steps should the U.S. Copyright Office take in this area? Is there any role for legislation?**

The Copyright Office could develop materials to help individual and small rightsholders to better understand when the DMCA safe harbors may be used (e.g., only for copyright claims, not for suppressing objectionable speech or for enforcing privacy preferences), and what information is necessary to provide in a valid takedown notice. In our experience, individual/small rightsholders often make mistakes about when the notice and takedown procedure is appropriate, and about what content is necessary to include. Anything that can be done to help educate such rightsholders would reduce the number of false or inadequate notices being sent, and improve the system.

**11. Several study participants pointed out that, since passage of the DMCA, no standard technical measures have been adopted pursuant to section 512(i). Should industry-wide or sub-industry- specific standard technical measures be adopted? If so, is there a role for government to help encourage the adoption of standard technical measures? Is legislative or other change required?**

The standard technical measures (“STM”) provision of Section 512(i) has failed. Section (i)(1)(B) mandates that any STM that is developed must be adopted or at least accommodated by all ISPs as a threshold eligibility requirement for the safe harbors. In this way, the statute mandates a one-size-fits-all technical approach to the problem of detecting and combating copyright infringement. Given the wide diversity of ISPs and types of content they host, this requirement is nonsensical. For example, if particular audio-processing software were developed pursuant to Section 512(i)(2)(A) and deemed to be a “STM” for the purposes of Section 512(i), even platforms that host only photographs or text documents would be required to run that audio processing software in order to be eligible for the safe harbor. Yet this would be an obvious waste of time and resources, as exactly zero copyright infringements could be prevented using this STM on such platforms.

This is a very good example of why policy makers should avoid mandating the creation of technologies that do not already exist at the time the regulation is drafted. See more on this topic below in our response to Question 12.

**12. Several study participants have proposed some version of a notice-and- stay-down system. Is such a system advisable? Please describe in specific detail how such a system should operate, and include potential legislative language, if appropriate. If it is not advisable, what particular problems would such a system impose? Are there ways to mitigate or avoid those problems? What implications, if any, would such a system have for future online innovation and content creation?**

Although we have very little detail about what a “notice and staydown” system might look like in reality, one thing is clear: it would create mandatory technical filtering of the Internet. This would be dangerous, unconstitutional, and technically infeasible.

The Internet Archive has been hailed as an “international treasure” for journalists because it is the only place reporters can consistently get access to materials that have been disappeared from the web.<sup>4</sup> There is no other non-profit organization dedicated to preserving the historical record of the Internet. Mandatory filtering would be uniquely problematic for us.

For one thing, the Internet Archive preserves the state of any given web page as *it existed on a particular date* via the Wayback Machine. Being forced to automatically remove material from the Wayback Machine would irreparably harm the historical record. This would be harmful for journalists who use the Wayback Machine to report on important stories of which there would be no evidence without the Archive. It would be harmful for attorneys and litigants who regularly use the Wayback Machine as evidence in legal proceedings. The very knowledge that a filter was running on the Wayback Machine would undermine its credibility as an accurate snapshot of the Internet at a given point in time. Therefore, filtering is a direct threat to our mission.

The Internet Archive also hosts the Political TV Ad Archive<sup>5</sup> and the TV News Archive.<sup>6</sup> As with the Wayback Machine, the very point of these archives is to preserve the historical record and ensure that politicians can be held accountable for their statements in ads or in TV appearances. A mandatory filter run on the TV News Archive might catch a famous song used in a political ad or at a campaign rally, and determine that such material must be removed. However, this would distort the historical record. This puts the Internet Archive in the untenable position of having to choose between protecting the historical record for future generations, and protecting its own legal interests.

Automated filters are not able to judge context. Even full-length copies of certain materials can be legal under certain circumstances. It could be fair use, or selective permission may have been granted by the rightsholder. But an algorithm will necessarily over-or-under block content because it is unable to detect these special circumstances. If an organization will be held legally liable for under-blocking, then of course it will choose to over-block, resulting in the censorship of legitimate online speech.

---

<sup>4</sup> *The Rachel Maddow Show* (MSNBC television broadcast, November 29, 2016) TV News Archive,

<https://archive.org/details/MSNBCRachelMaddow11292016WaybackMachine?start=267>

<sup>5</sup> Political TV Ad Archive, <https://politicaladarchive.org/>

<sup>6</sup> TV News Archive, <https://archive.org/details/tv>

Further, policing copyrighted content is a poor task to delegate to fully automated systems because the decisions affect people's fundamental expression rights. This is why the Internet Archive ensures that a real person reviews all DMCA takedown notices. However, the "human in the loop" for copyright removals really ought to represent and work for the rightsholder, not the intermediary. This is because (1) only the rightsholders know which uses are licensed (even they may not know, but they have a better chance than anyone else); and (2) as a matter of incentives, rightsholders have the liberty of erring on the side of leaving content online in ambiguous fair use calls, while the intermediary can only do so at risk of copyright damages.

And for all the rightsholder talk that the technology already exists, those of us who actually understand and develop technology can attest that developing a even remotely accurate filter that will work for each and every platform on the web and every type of content would be an extremely costly endeavor. YouTube famously spent \$60 million developing its Content ID system--which only works for audio and video content and rightsholders still complain about it all the time. It is very expensive and burdensome to create filters, and accurate, functional, cross-platform filters do not already exist. Nonprofits, libraries, and educational institutions who act as Internet service providers would be forced to spend a huge amount of their already scarce resources on policing copyright instead of focusing on their core missions.

There are many other objections to the notice and staydown concept, depending on how it could be implemented. For example, it could threaten user privacy and/or violate core due process rights. In the abstract, it is difficult to fully assess the kinds of damage that could ensue from such a system. For now, we therefore focus on the general objection that mandatory filtering of the Internet for copyright enforcement purposes is a very bad idea that has the potential to do far more harm than good.

**14. Several study participants mentioned concerns regarding certain case law interpretations of the existing provisions of section 512. Additionally, two new judicial decisions have come out since the first round of public comments was submitted in April 2016. What is the impact, if any, of these decisions on the effectiveness of section 512? If you believe it would be appropriate to address or clarify existing provisions of section 512, what would be the best ways to address such provisions (i.e., through the courts, Congress, the Copyright Office, and/or voluntary measures)? Please provide specific recommendations, such as legislative language, if appropriate.**

In *Capitol Records v. Vimeo*, the Second Circuit held that the DMCA safe harbors apply to pre-1972 sound recordings. This holding agrees with the statement of the Copyright Office that there is no policy justification for excluding older sound recordings from

Section 512.<sup>7</sup> We agree that this is the correct outcome, and it is good for libraries who seek to preserve older works. As we mentioned in our March 22, 2016 comments, part of our preservation work relies on community members and outside archivists who post digitized versions of older, at-risk materials, such as 78rpm records and cylinder recordings.<sup>8</sup> Many of these works were recorded prior to 1972. The physical condition of these sorts of works is extremely fragile. It is not likely that they will last until they become part of the public domain in 2067. Indeed, they will fall to dust well before they fall into the public domain. Many pre-1972 sound recordings are orphan works or works that have long since outlived their commercial lives, if they ever had one. Even so, because musical copyright is so complex, and since it is unclear whether fair use or other federal exceptions may apply to pre-1972 sound recordings, some libraries and other cultural institutions may determine that it is too risky to attempt to preserve these materials on their own. The DMCA safe harbor therefore provides an additional legal route to ensuring that these materials remain available to future generations.

We do not believe that any change is necessary to the statutory language, however, if any change is made it should be to clarify that the DMCA safe harbors apply to all copyrighted materials, not just those protected by federal copyright.

---

<sup>7</sup> Report of the Register of Copyrights, *Federal Protection for Pre-1972 Sound Recordings* (2011) at 130.

<sup>8</sup> See <https://archive.org/details/78rpm>.